Barclays Bank Case Study: Using Artificial Intelligence to Benchmark Organizational Data Flow Quality

Adrian McKeon
Infoshare Limited
amckeon@infoshare-is.com

Executive Summary/Abstract

Most IT systems cannot measure the accuracy of outputs
Does the system work and where is the evidence?
Are business decisions based on garbage?
Small data quality flaws at the start of a project magnify into inexplicable defects in end user outputs. Objectives cannot be translated into measurable performance indicators and nobody knows why.
Barclays1 used artificial intelligence to audit trail the history of each data record at sub field, field and record level from the source system to the warehouse. Audit trails made the workings of the IT system transparent and end users were able to validate output, identify errors and track back to fix them.


Which Comes First: Data or Technology?

Why measuring data quality is important
What happens when you can’t measure it
Barclays Bank Case Study
• What did Barclays Achieve?
• How did Barclays Achieve it?
Conclusions

Why Measuring Data Quality is Important

Data links policy to operations i.e. it translates objectives into measurable performance indicators
Output measurements reconcile silos of data from disparate agencies and link it to objectives
Decisions made on fragmented or inaccurate data are invariably bad making partnerships ineffective.
Analysis tools are useless if they do not provide accurate, measurable intelligence for action, monitoring and evaluation


The Problem: Measuring Data Quality Outputs

Knowledge Worker

Outputs (Analysis)

Accuracy ? Without evidence verification is just opinion

Legacy Systems

Enterprise Systems

Knowledge Worker

Extract Transform Load

Data Warehouse

Data Mart

CRM, ERP, Other

Other

Meta Data

Outputs (Analysis)
When things goes wrong...

Barclays Bank Case Study

What did Barclays Achieve?

- Problem
- Goals
- Requirements
- Results
- Other Applications which emerged

Barclays Goals & Requirement

The Problem

Goals
- Rationalise existing data maps
- Facilitate Basel II compliance through improved audit trails
- Maxime existing IT investment
- Maximise capital efficiency and RORA/ROE

Requirements
- Link millions of records from multiple systems in different countries
- Sub field audit trails to prove record accuracy
- Existing IT
  - Work with any system anywhere without disruption
  - No scrap/rework/new investment
  - Solution: turnkey, automated, no expertise to run
  - Intelligence created must prove ROI on existing IT

The Results

- Timely accurate, single view of customers available when required
- Solution is automated and requires little expertise to run
- Process sits alongside existing IT
- Detailed audit reports generated automatically to facilitate customer analysis at single record or aggregated record level
- Business rules can be changed to reflect emerging Basel requirements
- Ability to identify and eliminate duplicate records across all data sets from the Management Information
- Exception reports are generated where further investigation might be required

Barclays Bank Case Study

How did Barclays Achieve it?

Fixing warehouse or CRM
Proving ROI on existing investment
Basel II, Sarbanes-Oxley, IAS2005, Pigs
Risk reduction / what if scenario planning
Planning / implementing new IT
Balanced scorecards at individual level
Increased capital efficiency
Cross Selling
Customer profitability
KYC and money laundering
How did Barclays Achieve it?

**PROBLEM**
- IT infrastructure is a black box
- Few understand how it works
- No measure of output validity
- Is data held on the IT system good or bad?

**SOLUTION**
- Audit trail the complete history of each record
- Know what is happening at each part of process. IT is no longer a black box. Output is validated.

---

**Barriers to Effective Audit Trail Creation**
- Extract data from source systems
- Create data map of entity and business relationships at sub-field level
- Build in client business rule matching criteria
- Frequently integrate source system changes with model to maintain currency

**Barclays Approach**
- Links millions of records across multiple systems without disruption
- Turnkey, automated; fits any IT, easy to use in house
- Any number of data sources added as required
- Regularly updated and easily changed to meet new reporting needs
- Test existing IT outputs and design changes

---

**FROM THE START**
Legacy data... each action generates an audit trail point

**TO THE FINISH**
Enterprise data... the audit trail records how the data was fixed, why decisions were made, who made them and when

---

**Virtual Data Model**
A flat text file of record entity and business relationships across multiple source systems.
- Links proven by sub field, field & record audit trails. Can be used by any system to achieve a single view of the client
- All working components scripted in simple ASCII. Any application, whether Oracle, SAP or DB2, can store the model and access it with standard tools and protocols
- Performance of complex queries benefits from executing in memory and does not suffer from the read/write time delays that can be so expensive in data warehouses
The Role of Artificial Intelligence in Creating a Virtual Data Model

- Automating cleaning generates 50-80% accuracy; the floating 20-50% inaccuracy is the cause of data quality troubles affecting industry at the moment.
- To get 90-95% accuracy, you need to work with people i.e. subject specialists, case workers, data owners, et al and ask if you have the data in this system and add it to that one, what can it tell you and why does it tell you it in that way?
- Once you have done this, you are then in a position where you can build an automated script within your virtual model that mimics the way a specialist thinks and collaborates with peers in other departments.
- If you focus on just data you get ‘the data tells you this’. What you really need is a method of finding what you want from the data and that is what the virtual data model delivers.

Business Rules Matrix for Exploring the Potential of a Virtual Data Model to Answer End User Questions

Conclusions

- Which comes First - Data Quality or IT Design?
  - Data Quality
    - Then use the outputs to plan/tune IT
- Why are sub field audit trails important?
  - Hard Evidence
    - You can only prove output quality or establish a baseline for measuring ROI with hard evidence such as that provided by a sub field audit trail.
- What is the role of Artificial Intelligence?
  - Complexity
    - Traditional IT cannot cope ever changing complex variables linking records or the infinite range of end user queries.